

Unit 02, 03 실습

구글 드라이브 연동 & CSV 파일 데이터 읽어오기

```

1 #구글 드라이브와 연동
2 from google.colab import drive
3 drive.mount('/content/gdrive')
4 directory_path = '/content/gdrive/My Drive/Colab Notebooks/파이썬과데이터분석/'
5
6 # 자신의 구글 드라이브 directory_path 위치에 seoul.csv 파일 저장

```

Mounted at /content/gdrive

csv.reader() : CSV 파일에서 데이터를 읽어오는 함수

csv.writer() : CSV 파일에 데이터를 저장하는 함수

헤더(header): 데이터의 속성을 설명, 첫번째 줄에 위치

저장이 완료되었습니다.



며 데이터의 탐색 위치를 다음 행으로 이동시키는 함수

```

1 import csv
2 #seoul.csv 데이터 한 행(row)씩 읽어오기
3 f = open(directory_path + 'seoul.csv', 'r', encoding='cp949')
4 data = csv.reader(f, delimiter=',')
5 #next() 함수를 활용해 헤더 저장하기
6 header = next(data)
7 for row in data :
8     print(row)

```

스트리밍 출력 내용이 길어서 마지막 5000줄이 삭제되었습니다.

```

['2004-07-20', '108', '25.7', '24.8', '26.8']
['2004-07-21', '108', '26.8', '24.1', '31.8']
['2004-07-22', '108', '27.4', '24.1', '32']
['2004-07-23', '108', '28.3', '24', '33.2']
['2004-07-24', '108', '27.6', '23.7', '30.8']
['2004-07-25', '108', '25.4', '22.6', '28']
['2004-07-26', '108', '26.5', '23.8', '29.2']
['2004-07-27', '108', '27.6', '24.7', '30.2']
['2004-07-28', '108', '28.7', '26', '32.7']
['2004-07-29', '108', '28.9', '24.1', '33.2']
['2004-07-30', '108', '29.2', '25.2', '33.4']
['2004-07-31', '108', '29.9', '24.7', '34.5']
['2004-08-01', '108', '29.4', '25.3', '34.2']
['2004-08-02', '108', '28.3', '23.7', '33.2']
['2004-08-03', '108', '28.8', '25.7', '32.7']
['2004-08-04', '108', '28.5', '23.2', '32.9']
['2004-08-05', '108', '27.6', '23.8', '32.7']
['2004-08-06', '108', '28.3', '24.7', '32.9']
['2004-08-07', '108', '28.5', '26.2', '32.4']
['2004-08-08', '108', '27.8', '26.1', '32']
['2004-08-09', '108', '28.6', '25', '33']

```

```
[ '2004-08-10', '108', '30.2', '25.2', '36.2' ]
[ '2004-08-11', '108', '30.4', '25.6', '35.7' ]
[ '2004-08-12', '108', '29.6', '25.6', '34.7' ]
[ '2004-08-13', '108', '28.7', '26', '32.7' ]
[ '2004-08-14', '108', '26.3', '24', '29.2' ]
[ '2004-08-15', '108', '26.7', '22.8', '31' ]
[ '2004-08-16', '108', '25', '21.5', '29.1' ]
[ '2004-08-17', '108', '23.1', '21.1', '26.2' ]
[ '2004-08-18', '108', '23.2', '22', '25' ]
[ '2004-08-19', '108', '21.8', '19.1', '25.3' ]
[ '2004-08-20', '108', '23.5', '20.5', '26.8' ]
[ '2004-08-21', '108', '23.3', '21', '25.5' ]
[ '2004-08-22', '108', '22.8', '21.3', '25.2' ]
[ '2004-08-23', '108', '22.8', '20.5', '26' ]
[ '2004-08-24', '108', '23.9', '18.7', '29.2' ]
[ '2004-08-25', '108', '24.2', '20.7', '27.5' ]
[ '2004-08-26', '108', '23.7', '21.2', '28.5' ]
[ '2004-08-27', '108', '24.6', '21.1', '28.3' ]
[ '2004-08-28', '108', '23.9', '22.1', '26.6' ]
[ '2004-08-29', '108', '25.5', '21.6', '30.1' ]
[ '2004-08-30', '108', '26', '22.7', '31.5' ]
[ '2004-08-31', '108', '24.6', '20', '29.3' ]
[ '2004-09-01', '108', '24.7', '22.2', '29.3' ]
[ '2004-09-02', '108', '24.1', '21.1', '29.2' ]
[ '2004-09-03', '108', '23.7', '19', '28' ]
[ '2004-09-04', '108', '24.9', '20.7', '30' ]
[ '2004-09-05', '108', '27', '31.3' ]
[ '2004-09-06', '108', '29.7' ]
[ '2004-09-07', '108', '24.2' ]
[ '2004-09-08', '108', '22.1', '18', '27.2' ]
[ '2004-09-09', '108', '23.2', '18.7', '28.8' ]
[ '2004-09-10', '108', '22.1', '17.6', '26.7' ]
[ '2004-09-11', '108', '17.4', '16', '20.6' ]
[ '2004-09-12', '108', '20.5', '16.6', '23.5' ]
[ '2004-09-13', '108', '21.4', '19.3', '24.3' ]
[ '2004-09-14', '108', '23', '17.5', '28.7' ]
[ '2004-09-15', '108', '24.5', '19.6', '28.2' ]
[ '2004-09-16', '108', '22.1', '18.7', '27.2' ]
```

저장이 완료되었습니다.



▼ 실습1:

질문 : 기상 관측 이래, 서울의 최고 기온이 가장 높았던 날은 언제였고, 몇 도였을까?

질문을 해결하는데 필요한 절차(알고리즘)는?

- 1 | 데이터를 읽어온다.
- 2 | 순차적으로 최고 기온을 확인한다.
- 3 | 최고 기온이 가장 높았던 날짜의 데이터를 저장한다.
- 4 | 최종 저장된 데이터를 출력한다.

```
1 #step 1) 데이터 불러서 한 행씩 출력하기
2 #step 2) 데이터 중 최고 기온을 실수로 변환하여 한 행씩 출력하기
3
```

```

4 f = open(directory_path + 'seoul.csv', 'r', encoding='cp949')
5 data = csv.reader(f)
6
7 header = next(data)
8 for row in data :
9     #누락된 데이터가 있으면 오류 발생
10    if row[-1] == '' :
11        row[-1] = -999 # -999를 넣어 빈 문자열이 있던 자리라고 표시
12    row[-1] = float(row[-1]) # 최고 기온을 실수로 변형
13    print(row[-1])
14
15 f.close()

```

```

1 #step 3) 최고 기온과 최고 기온이었던 날짜 찾기
2
3 max_temp = -999 # 최고 기온 값을 저장할 변수
4 max_date = '' # 최고 기온이 가장 높았던 날짜를 저장할 변수
5 f = open(directory_path + 'seoul.csv', 'r', encoding='cp949')
6 data = csv.reader(f)
7 header = next(data)
8
9 for row in data :
10    if row[-1] == '' :

```

저장이 완료되었습니다.



```

13    if max_temp < row[-1] :
14        max_date = row[0]
15        max_temp = row[-1]
16
17 f.close()
18 print('기상 관측 이래 서울의 최고 기온이 가장 높았던 날은', max_date + '로, ', max_temp, '도 였습L
19

```

기상 관측 이래 서울의 최고 기온이 가장 높았던 날은 1994-07-24로, 38.4 도 였습니다.

Unit 04, 05 실습

matplotlib: 2D 형태의 그래프, 이미지 등을 그릴 때 사용

실습1: 기본 그래프 그리기

- plot: 꺾은선 그래프 그리기

```

1 #matplotlib 라이브러리의 pyplot 모듈을 'plt' 라는 별명으로 부르기(as, alias)
2 import matplotlib.pyplot as plt
3

```

```

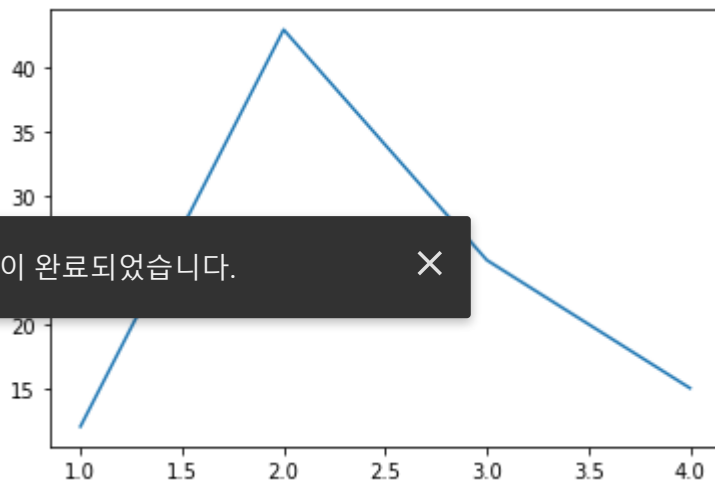
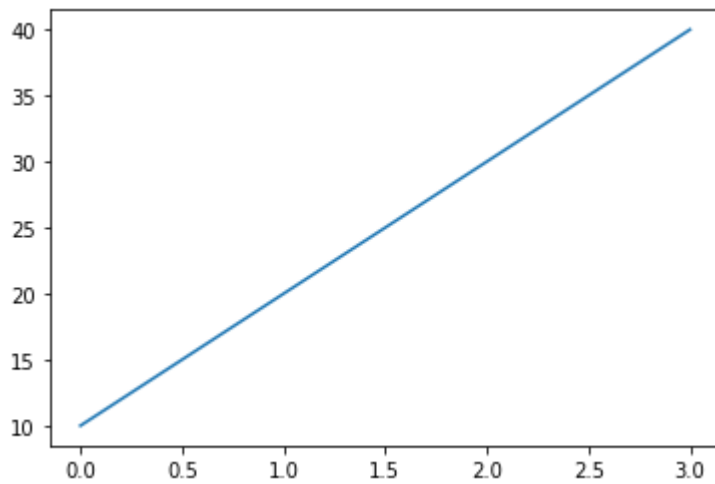
1
2 plt.plot([10, 20, 30, 40])
3 plt.show()

```

```

3 plt.show()
4
5 plt.plot([1,2,3,4], [12, 43, 25, 15])
6 plt.show()

```

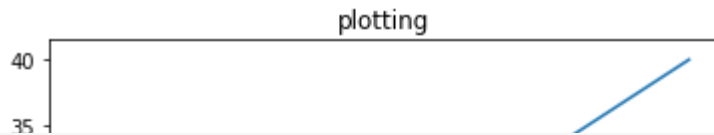


실습 2: 그래프에 옵션 추가하기

```

1 #그래프에 제목 넣기
2 '''
3 code 추가
4 '''
5 plt.title('plotting')
6 plt.plot([10, 20, 30, 40])
7 plt.show()
8

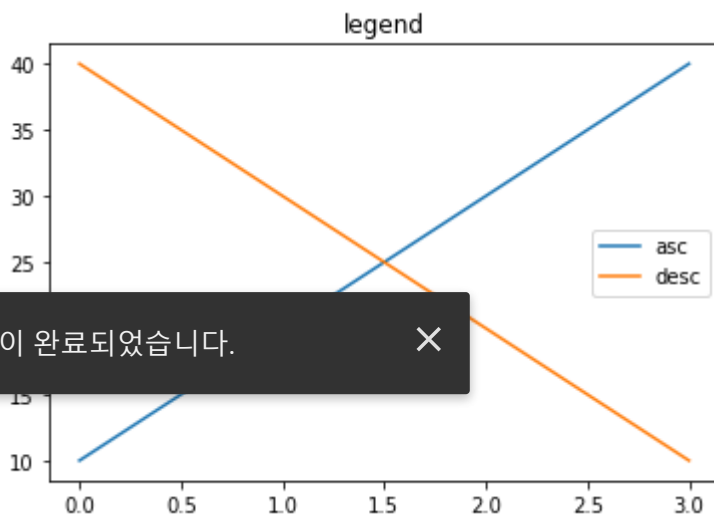
```



```

1  #그래프에 범례(기호 설명표) 넣기
2  '''
3  code 추가
4  '''
5  plt.title('legend')
6  plt.plot([10, 20, 30, 40], label='asc') # 증가를 의미하는 asc 범례
7  plt.plot([40, 30, 20, 10], label='desc') # 감소를 의미하는 desc 범례
8  plt.legend()
9  plt.show()
10

```



```

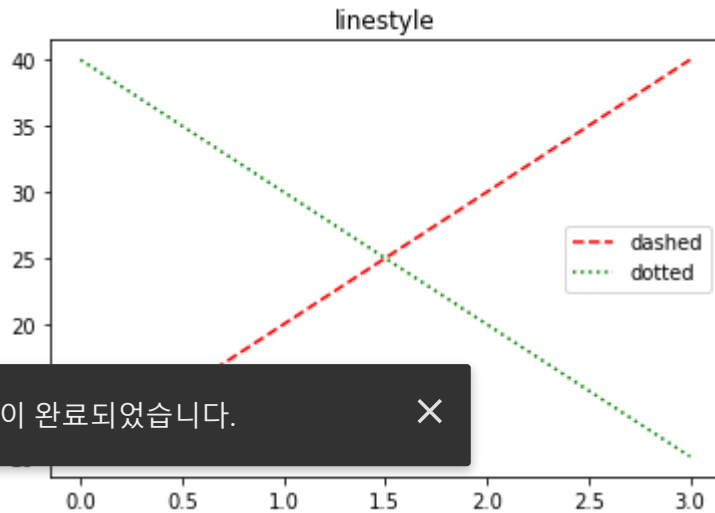
1  #그래프 색상 바꾸기
2  '''
3  code 추가
4  '''
5  plt.title('color') # 제목 설정
6  plt.plot([10, 20, 30, 40], color='skyblue', label='skyblue') # 하늘색 그래프
7  plt.plot([40, 30, 20, 10], 'pink', label='pink') # 분홍색 그래프
8  plt.legend() # 범례 표시
9  plt.show()
10

```

```

1 #그래프 선 모양 바꾸기
2 '''
3 code 추가
4 '''
5 plt.title('linestyle') #제목 설정
6 plt.plot([10, 20, 30, 40], color='r', linestyle='--', label='dashed') # 빨간색 dashed 그래프
7 plt.plot([40, 30, 20, 10], color='g', ls=':', label='dotted') # 초록색 dotted 그래프
8 plt.legend() # 범례 표시
9 plt.show()
10

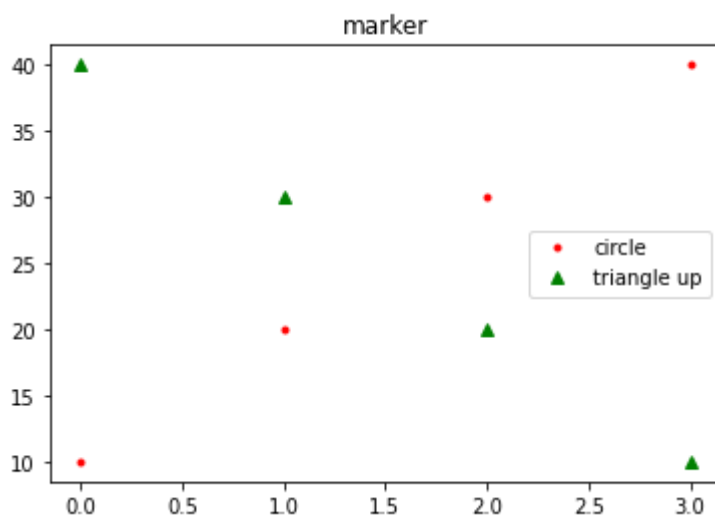
```



```

1 # 그래프 마커 모양 바꾸기
2 plt.title('marker') # 제목 설정
3 '''
4 code 추가
5 '''
6 plt.plot([10, 20, 30, 40], 'r.', label='circle') # 빨간색 원형 마커 그래프
7 plt.plot([40, 30, 20, 10], 'g^', label='triangle up') # 초록색 삼각형 마커 그래프
8 plt.legend() # 범례 표시
9 plt.show()

```

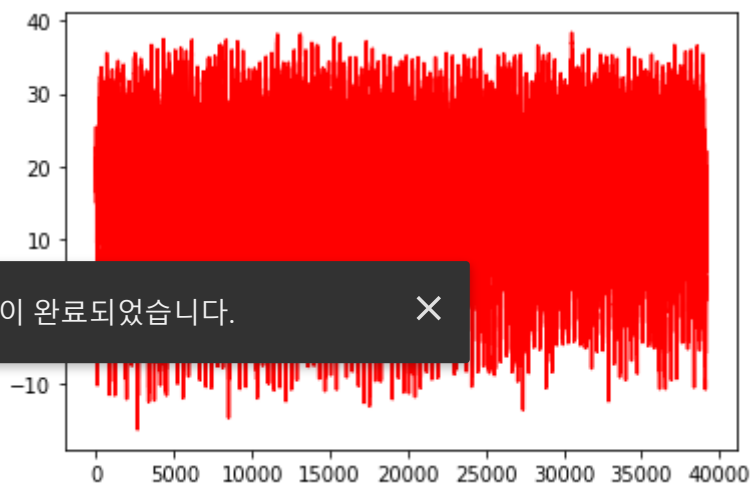


내 생일의 기온 변화를 그래프로 그리기

```

1  import csv
2  import matplotlib.pyplot as plt
3  f = open(directory_path + 'seoul.csv', 'r', encoding='cp949')
4  data = csv.reader(f)
5  next(data)
6  result = []
7
8  for row in data :
9      if row[-1] != '' :
10         result.append(float(row[-1]))
11
12  plt.plot(result, 'r') # result 리스트에 저장된 값을 빨간색 그래프로 그리기
13  plt.show() # 그래프 나타내기

```



실습3: 날짜 데이터 추출하기

```

1
2  date = '1907-10-01'
3  print(date.split('-'))
4
5  print(date.split('-')[0])
6  print(date.split('-')[1])
7  print(date.split('-')[2])
8

```

```

['1907', '10', '01']
1907
10
01

```

실습4: 내 생일의 최고 기온 및 최저 기온 데이터 시각화하기

생일: 2월 14일

구글 코랩 한글 사용

1. 폰트 설치

2. 런타임 - 런타임 다시 시작 을 클릭하여 런타임을 재시작

```

1 # 폰트 설치
2 !sudo apt-get install -y fonts-nanum
3 !sudo fc-cache -fv
4 !rm ~/.cache/matplotlib -rf

```

```

Reading package lists... Done
Building dependency tree
Reading state information... Done
The following package was automatically installed and is no longer required:
  libnvidia-common-460
Use 'sudo apt autoremove' to remove it.
The following NEW packages will be installed:
  fonts-nanum
0 upgraded, 1 newly installed, 0 to remove and 34 not upgraded.
Need to get 9,604 kB of archives.
After this operation, 29.5 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu bionic/universe amd64 fonts-nanum all 20170925-1 [9,604 kB]
Fetched 9,604 kB in 1s (8,257 kB/s)
debconf: unable to initialize frontend: Dialog
debconf: (No usable dialog-like program is installed, so the dialog based frontend cannot be used by default. Falling back to readline)
debconf: falling back to frontend: Readline
debconf: (The frontend controlling tty.)
debconf: falling back to frontend: Teletype
dpkg-preconfigure: unable to re-open stdin:
Selecting previously unselected package fonts-nanum.
(Reading database ... 160690 files and directories currently installed.)
Preparing to unpack .../fonts-nanum_20170925-1_all.deb ...
Unpacking fonts-nanum (20170925-1) ...
Setting up fonts-nanum (20170925-1) ...
Processing triggers for fontconfig (2.12.6-0ubuntu2) ...
/usr/share/fonts: caching, new cache contents: 0 fonts, 1 dirs
/usr/share/fonts/truetype: caching, new cache contents: 0 fonts, 3 dirs
/usr/share/fonts/truetype/humor-sans: caching, new cache contents: 1 fonts, 0 dirs
/usr/share/fonts/truetype/liberation: caching, new cache contents: 16 fonts, 0 dirs
/usr/share/fonts/truetype/nanum: caching, new cache contents: 10 fonts, 0 dirs
/usr/local/share/fonts: caching, new cache contents: 0 fonts, 0 dirs
/root/.local/share/fonts: skipping, no such directory
/root/.fonts: skipping, no such directory
/var/cache/fontconfig: cleaning cache directory
/root/.cache/fontconfig: not cleaning non-existent cache directory
/root/.fontconfig: not cleaning non-existent cache directory
fc-cache: succeeded

```

저장이 완료되었습니다.



```

1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open(directory_path + 'seoul.csv', 'r', encoding='cp949')
5 data = csv.reader(f)
6 next(data)
7 high = [] # 최고 기온 값을 저장할 리스트 high 생성
8 low = [] # 최저 기온 값을 저장할 리스트 low 생성
9

```



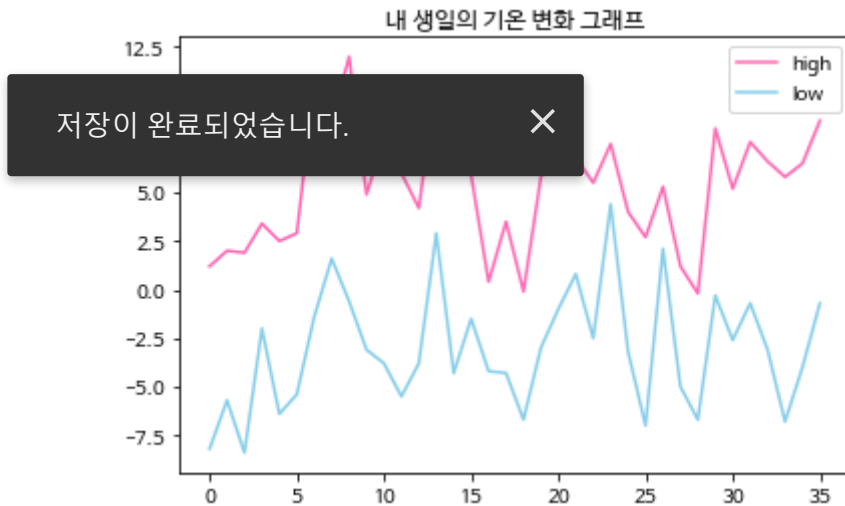
```

10  '''
11  code 추가
12  '''
13  for row in data :
14      if row[-1] != '' and row[-2] != '' :
15          date = row[0].split('-')
16          if 1983 <= int(date[0]) :
17              if date[1] == '02' and date[2] == '14' :
18                  high.append(float(row[-1]))
19                  low.append(float(row[-2]))
20
21  plt.rc('font', family='NanumBarunGothic')
22  #plt.rc('font', family = 'Malgun Gothic')
23  plt.rcParams['axes.unicode_minus'] = False
24  plt.title('내 생일의 기온 변화 그래프')
25  plt.plot(high, 'hotpink', label = 'high')
26  plt.plot(low, 'skyblue', label = 'low')
27  plt.legend()
28  plt.show()
29

```

최고 기온 값과 최저 기온 값이 존재한
날짜 값을 - 문자를 기준으로 구분하여
1983년 이후 데이터라면
2월 14일이라면
최고 기온 값을 high 리스트에 저장
최저 기온 값을 low 리스트에 저장

나눔 고딕을 기본 글꼴로 설정
맑은 고딕을 기본 글꼴로 설정
마이너스 기호 깨짐 방지
제목 설정
high 리스트에 저장된 값을 hotpink 색
low 리스트에 저장된 값을 skyblue 색
범례 표시
그래프 나타내기



===== 실습 끝 =====



저장이 완료되었습니다. 